

# Évaluation de la qualité des données

Conduire une évaluation de la qualité des données de routine comme le ferait un auditeur bailleur — facteurs de vérification, cohérence temporelle, et un rapport qui nomme la correction autant que le défaut.

|                         |                                                                                                                     |
|-------------------------|---------------------------------------------------------------------------------------------------------------------|
| Formateur               | Pierre Robentz Cassion                                                                                              |
| Niveau                  | Intermédiaire                                                                                                       |
| Heures apprenant        | 12 h                                                                                                                |
| Enseignée en            | Python · R                                                                                                          |
| Mise à jour             | 28 juillet 2026                                                                                                     |
| Secteurs                | Santé publique · Nutrition                                                                                          |
| Normes et méthodologies | Définitions d'indicateurs de l'UNICEF · Norme humanitaire fondamentale (CHS) · Gestion axée sur les résultats (GAR) |

Consulter et exécuter la formation en ligne sur [data-analysis.cassion.dev/fr/cours/data-quality-assessment](https://data-analysis.cassion.dev/fr/cours/data-quality-assessment)

Le contenu pédagogique est diffusé sous licence CC BY 4.0. Vous pouvez l'adapter et le rediffuser, y compris à des fins commerciales, en créditant Cassion.

## Ce que vous saurez faire

- Noter un jeu de données sur les cinq dimensions de la qualité, chacune avec sa mesure, au lieu de le déclarer « bon » ou « mauvais »
- Calculer la complétude et la promptitude du rapportage comme des indicateurs à part entière, avec leurs propres dénominateurs
- Recompter un registre source face à ce qui a été remonté et transformer l'écart en facteur de vérification assorti d'une bande de tolérance
- Choisir les formations sanitaires à vérifier lorsqu'on ne peut pas toutes les visiter, et énoncer ce que l'échantillon permet ou non de conclure
- Détecter la préférence de chiffres, l'attraction des valeurs rondes et les séries anormalement stables, et étalonner le contrôle sur ce que le hasard produit
- Rédiger un rapport d'évaluation dont chaque constat porte une cause racine, une action corrective, un responsable et une échéance

## Prérequis

Aucun. Cette formation ne suppose aucune expérience préalable de la programmation.

# Sommaire

|           |                                                                       |           |
|-----------|-----------------------------------------------------------------------|-----------|
| <b>I</b>  | <b>Ce que qualité veut dire ici</b>                                   | <b>1</b>  |
| <b>1</b>  | <b>Cinq dimensions, cinq mesures</b>                                  | <b>2</b>  |
| 1.1       | La phrase qu'il faut cesser de dire . . . . .                         | 2         |
| 1.2       | Les cinq, et la mesure de chacune . . . . .                           | 2         |
| 1.3       | Calculez les cinq sur une seule extraction . . . . .                  | 2         |
| 1.4       | Pondérez les dimensions par ce qu'elles coûtent . . . . .             | 3         |
| 1.5       | Ne réduisez pas à un seul nombre . . . . .                            | 4         |
| 1.6       | La suite . . . . .                                                    | 4         |
| <b>2</b>  | <b>Les deux indicateurs que personne ne calcule</b>                   | <b>5</b>  |
| 2.1       | Deux indicateurs, et aucun n'est dans le rapport . . . . .            | 5         |
| 2.2       | La complétude du rapportage . . . . .                                 | 5         |
| 2.3       | Décomposez toujours de trois façons . . . . .                         | 6         |
| 2.4       | Le piège du dénominateur . . . . .                                    | 6         |
| 2.5       | La promptitude exige un champ que vous n'avez peut-être pas . . . . . | 7         |
| 2.6       | Lisez tous les autres chiffres à travers ces deux-là . . . . .        | 8         |
| 2.7       | La suite . . . . .                                                    | 8         |
| <b>II</b> | <b>Confronter à la source</b>                                         | <b>9</b>  |
| <b>3</b>  | <b>Le recomptage, et le nombre qu'il produit</b>                      | <b>10</b> |
| 3.1       | L'exactitude est la dimension qui exige un déplacement . . . . .      | 10        |
| 3.2       | Le facteur de vérification . . . . .                                  | 10        |
| 3.3       | Calculez-le sur un registre dont vous disposez . . . . .              | 10        |
| 3.4       | Le chiffre du district est le piège . . . . .                         | 11        |
| 3.5       | La bande de tolérance . . . . .                                       | 12        |
| 3.6       | Le recomptage est un protocole, pas une après-midi . . . . .          | 12        |
| 3.7       | Ce qu'un facteur de vérification ne peut pas dire . . . . .           | 13        |
| 3.8       | La suite . . . . .                                                    | 13        |
| <b>4</b>  | <b>Choisir les six formations sanitaires à visiter</b>                | <b>14</b> |
| 4.1       | Vous avez une semaine et trente-huit formations sanitaires . . . . .  | 14        |
| 4.2       | Aléatoire — la seule qui se généralise . . . . .                      | 14        |
| 4.3       | Fondée sur le risque — là où sont les problèmes . . . . .             | 15        |
| 4.4       | Recensement des plus grosses . . . . .                                | 15        |
| 4.5       | Combien suffisent . . . . .                                           | 16        |
| 4.6       | LQAS — conçu exactement pour cela . . . . .                           | 16        |
| 4.7       | Combinez, et étiquetez . . . . .                                      | 16        |
| 4.8       | La suite . . . . .                                                    | 17        |

|            |                                                                             |           |
|------------|-----------------------------------------------------------------------------|-----------|
| <b>III</b> | <b>Repérer les sites qui méritent une visite</b>                            | <b>18</b> |
| <b>5</b>   | <b>Comparer une formation sanitaire à son propre passé</b>                  | <b>19</b> |
| 5.1        | La cohérence est une comparaison, et vous en choisissez le terme . . . . .  | 19        |
| 5.2        | Rapport à la médiane . . . . .                                              | 19        |
| 5.3        | Le problème des petits effectifs . . . . .                                  | 20        |
| 5.4        | Adjoignez un plancher absolu au rapport . . . . .                           | 20        |
| 5.5        | Le sens n'est pas symétrique . . . . .                                      | 21        |
| 5.6        | Des séries trop stables . . . . .                                           | 21        |
| 5.7        | Classez les formations, ne signalez pas des lignes . . . . .                | 22        |
| 5.8        | La suite . . . . .                                                          | 22        |
| <b>6</b>   | <b>Ce que dit le dernier chiffre</b>                                        | <b>23</b> |
| 6.1        | Les mesures ont une texture . . . . .                                       | 23        |
| 6.2        | Le contrôle . . . . .                                                       | 23        |
| 6.3        | L'attraction des âges, et pourquoi ce n'est la faute de personne . . . . .  | 24        |
| 6.4        | Étalonnez sur ce que produit le hasard . . . . .                            | 24        |
| 6.5        | Pourquoi le cas négatif compte davantage . . . . .                          | 25        |
| 6.6        | Ce que vous avez le droit d'écrire . . . . .                                | 26        |
| 6.7        | La suite . . . . .                                                          | 26        |
| <b>IV</b>  | <b>Transformer les constats en changement</b>                               | <b>27</b> |
| <b>7</b>   | <b>Du défaut à ce qui l'a produit</b>                                       | <b>28</b> |
| 7.1        | Un constat n'est pas une cause, et seule une cause se corrige . . . . .     | 28        |
| 7.2        | Cinq catégories, et elles n'exigent pas le même argent . . . . .            | 28        |
| 7.3        | Laissez les données réduire le champ avant d'interroger quiconque . . . . . | 28        |
| 7.4        | Puis interrogez le formulaire . . . . .                                     | 29        |
| 7.5        | Les cinq pourquoi, avec une règle d'arrêt . . . . .                         | 29        |
| 7.6        | Testez la cause sur les données . . . . .                                   | 30        |
| 7.7        | Placez la cause au niveau où vit le remède . . . . .                        | 30        |
| 7.8        | La suite . . . . .                                                          | 31        |
| <b>8</b>   | <b>Le rapport qui change le tour suivant</b>                                | <b>32</b> |
| 8.1        | Qui le lit réellement . . . . .                                             | 32        |
| 8.2        | L'ordre . . . . .                                                           | 32        |
| 8.3        | Le tableau des constats . . . . .                                           | 32        |
| 8.4        | Des constats qu'un lecteur peut vérifier . . . . .                          | 33        |
| 8.5        | Notez les dimensions, ne notez pas le district . . . . .                    | 33        |
| 8.6        | Le tour de suivi est l'essentiel . . . . .                                  | 34        |
| 8.7        | La présenter à ceux qu'elle décrit . . . . .                                | 35        |
| 8.8        | Ce que ce cours, et ce module, vous laissent . . . . .                      | 35        |

## À propos de cette formation

Tout le monde dans ce secteur a assisté à la réunion où quelqu'un déclare que la qualité des données est mauvaise. Cette phrase n'est jamais utile. Mauvaise en quoi — incomplète, tardive, incohérente avec le registre, ou contradictoire avec elle-même ? Mauvaise partout, ou dans quatre formations sanitaires sur trente-huit ? Assez mauvaise pour changer la décision, ou mauvaise d'une manière qui déplace le chiffre d'un demi-point ?

Une évaluation de la qualité des données est la discipline qui remplace cette phrase par des nombres. Elle a une forme sur laquelle le secteur s'est entendu — **cinq dimensions, chacune avec sa mesure ; un recomptage face à la source ; un échantillon défendable ; et un rapport où chaque constat a un responsable et une date**. Les bailleurs la conduisent sur vous, et la conduire sur vous-même d'abord fait la différence entre un constat et une surprise.

Ce cours achève le module. *Nettoyage et validation des données* vous a appris à trouver et corriger les défauts d'un fichier. *Jointures et restructuration* vous a appris à assembler des fichiers et à les rapprocher. Celui-ci transforme les deux en une routine qu'un tiers peut exécuter selon un calendrier — et en un document qui change ce qui se passera au trimestre suivant, seule raison de faire tout cela.

Chaque technique est présentée en Python et en R. Les exemples s'exécutent sur l'extraction vaccinale de type DHIS2, où 642 des 2 736 couples formation-mois ne portent aucun rapport, et sur l'enquête SMART, où les mesures de taille d'une équipe se terminent par ,0 ou ,5 dans soixante-neuf pour cent des cas.

Un avertissement sur le ton, car il décide de l'utilité de ce travail. Une évaluation qui se lit comme une accusation produit un rapportage défensif, et un rapportage défensif produit de moins bonnes données. **Le constat porte sur un système, pas sur une personne**, et le cours revient régulièrement sur la manière de l'écrire.

## Première partie

# Ce que qualité veut dire ici

## CHAPITRE 1

# Cinq dimensions, cinq mesures

Leçon 1 sur 8 · 75 min

« *Les données sont mauvaises* » n'est pas un constat. *Exactitude, complétude, actualité, cohérence et intégrité* — chacune avec un nombre, calculée sur la même extraction, et un tableau de bord qui résiste à la contestation.

### 1.1 La phrase qu'il faut cesser de dire

« La qualité des données est mauvaise. » Tout le monde acquiesce, rien ne se passe, et la même phrase revient au trimestre suivant.

Elle échoue parce qu'elle est inexploitable de trois façons distinctes. Elle ne dit pas **quelle** qualité — un rapport tardif et un rapport fabriqué sont deux problèmes aux remèdes différents. Elle ne dit pas **combien** — mauvaise partout ou mauvaise dans quatre formations sanitaires. Et elle ne dit pas **et alors** — si le défaut est assez grand pour changer le chiffre sur lequel quelqu'un va agir.

Les cinq dimensions existent pour corriger les trois d'un coup. Ce n'est pas un cadre à réciter, ce sont cinq questions dont la réponse est un nombre.

### 1.2 Les cinq, et la mesure de chacune

| Dimension         | La question                                                     | La mesure                               |
|-------------------|-----------------------------------------------------------------|-----------------------------------------|
| <b>Complétude</b> | Tout le monde a-t-il rapporté, et sur tout ?                    | Taux de rapportage ; taux de valeur     |
| <b>Actualité</b>  | Le rapport est-il arrivé à temps pour servir ?                  | Part transmise avant l'échéance ; ret   |
| <b>Exactitude</b> | Le chiffre remonté correspond-il à la source ?                  | Facteur de vérification — recompta      |
| <b>Cohérence</b>  | Les nombres s'accordent-ils entre eux et avec le mois dernier ? | Taux d'échec des règles internes ; ta   |
| <b>Intégrité</b>  | La valeur vient-elle d'une mesure ou d'autre chose ?            | Préférence de chiffres ; attraction ; s |

Deux points de ce tableau méritent discussion, car les deux reviennent toujours.

**L'exactitude est la seule qui exige le document source.** Les quatre autres se calculent sur l'extraction que vous avez déjà, à votre bureau, cet après-midi. Cette asymétrie commande toute la conception d'une évaluation — les quatre dimensions de bureau servent à décider où dépenser la cinquième, la coûteuse.

**L'intégrité n'est pas une accusation.** Elle mesure si les nombres ressemblent à des mesures. Une distribution d'âges attirée par les valeurs rondes signifie généralement que personne n'a demandé d'acte de naissance, non que quelqu'un a inventé quoi que ce soit, et la leçon qui lui est consacrée passe l'essentiel de son temps sur cette distinction.

### 1.3 Calculez les cinq sur une seule extraction

Prenez l'extraction vaccinale de routine — 38 formations sanitaires, douze mois, six antigènes.

```
import pandas as pd
```

```

vax = pd.read_csv("vaccination-coverage-2024.v1.csv", parse_dates=["period"])
vax["reported"] = vax["report_submitted"] == True

scorecard = {}

# Completeness
scorecard["reporting_rate"] = vax["reported"].mean()
scorecard["field_completeness"] = 1 - vax.isna().mean().mean()

# Consistency: doses cannot exceed the target population
rule_fail = (vax["doses_administered"] > vax["target_population"]).mean()
scorecard["rule_failure_rate"] = rule_fail

# Integrity: share of reported values ending in zero
reported = vax[vax["reported"]]
scorecard["round_number_share"] = (reported["doses_administered"] % 10 == 0).mean()

print(pd.Series(scorecard).round(3))

library(dplyr)
library(readr)

vax <- read_csv("vaccination-coverage-2024.v1.csv")

scorecard <- tibble::tibble(
  reporting_rate      = mean(vax$report_submitted),
  field_completeness = 1 - mean(is.na(as.matrix(vax))),
  rule_failure_rate   = mean(vax$doses_administered > vax$target_population),
  round_number_share = mean(
    vax$doses_administered[vax$report_submitted] %% 10 == 0
  )
)

round(scorecard, 3)

```

| Mesure                  | Valeur | À lire comme                                                                 |
|-------------------------|--------|------------------------------------------------------------------------------|
| Taux de rapportage      | 76,5 % | 642 des 2 736 couples formation-mois ne portent aucun rapport                |
| Complétude des champs   | 100 %  | aucune case vide — ce qui n'est pas la même chose qu'aucune donnée manquante |
| Taux d'échec des règles | 0 %    | aucune formation ne rapporte plus de doses que sa population cible           |
| Part de nombres ronds   | 10,4 % | ce que produit le hasard ; rien à voir                                       |

Lisez les lignes deux et un ensemble. **La complétude des champs vaut 100 % et la complétude du rapportage 76,5 %**, et seule la seconde compte. Chaque couple formation-mois sans rapport figure dans le fichier comme une ligne de zéros avec un indicateur, si bien qu'un décompte de cases vides déclare les données parfaites. C'est le défaut autour duquel tout le module tourne — un rapport manquant qui arrive sous forme de zéro.

## 1.4 Pondérez les dimensions par ce qu'elles coûtent

Un tableau de bord à cinq nombres égaux laisse entendre que les cinq comptent également. Ce n'est pas le cas, et laquelle prime dépend entièrement de l'usage des données.

- Pour un **taux de couverture**, la complétude domine. Un taux de rapportage de 76,5 % fait de tout total de district une borne inférieure, et aucune exactitude dans les rapports reçus n'y remédie.

- Pour une **charge de cas servant à approvisionner**, l'exactitude domine. Commander des aliments thérapeutiques sur une charge surestimée gaspille de l'argent ; sur une charge sous-estimée, des enfants n'en reçoivent pas.
- Pour un **indicateur d'alerte précoce**, l'actualité domine. Un rapport parfaitement exact arrivant six semaines après l'épidémie ne vaut rien.

**Dites quelle dimension vous avez pondérée et pourquoi, dans le rapport.** Un score composite qui cache sa pondération est le même défaut qu'un indicateur qui cache son dénominateur.

## 1.5 Ne réduisez pas à un seul nombre

Quelqu'un demandera un score unique sur 100. Résistez, et proposez le tableau de bord à la place, pour une raison précise — **les cinq dimensions ont des remèdes différents, et le nombre unique ne dit rien sur celui à appliquer.**

Un district à 82 % parce qu'il est en retard se corrige en déplaçant une échéance. Un district à 82 % parce qu'un tiers de ses formations ne rapportent jamais se corrige en cherchant pourquoi elles ont cessé. Les deux font 82. Un seul se corrige par un atelier de formation, et aucun ne se corrige par l'atelier que quelqu'un va proposer.

Si un composite est réellement exigé — certains modèles bailleurs l'imposent — publiez-le **à côté** de ses composantes, jamais à leur place.

```
scorecard_table = pd.DataFrame({
    "dimension": ["Completeness", "Timeliness", "Accuracy", "Consistency", "Integrity"],
    "measure": ["Reporting rate", "Submitted by deadline", "Verification factor",
               "Rule failure rate", "Round-number share"],
    "value": [0.765, None, None, 0.0, 0.104],
    "source": ["extract", "submission log", "facility visit", "extract", "extract"],
})
print(scorecard_table)
```

```
scorecard_table <- tibble::tribble(
  ~dimension, ~measure, ~value, ~source,
  "Completeness", "Reporting rate", 0.765, "extract",
  "Timeliness", "Submitted by deadline", NA, "submission log",
  "Accuracy", "Verification factor", NA, "facility visit",
  "Consistency", "Rule failure rate", 0.000, "extract",
  "Integrity", "Round-number share", 0.104, "extract"
)
```

**C'est la colonne source qui est utile.** Elle dit immédiatement quels nombres vous pouvez produire aujourd'hui et lesquels exigent un déplacement, et c'est la forme de tout plan d'évaluation.

## 1.6 La suite

Deux des cinq dimensions ne sont presque jamais rapportées, et toutes deux se calculent sur des données que vous détenez déjà. La leçon suivante les calcule correctement — complétude et promptitude du rapportage, chacune avec son propre dénominateur, et le piège consistant à diviser par les formations qui ont rapporté.

## CHAPITRE 2

# Les deux indicateurs que personne ne calcule

Leçon 2 sur 8 · 75 min

*Complétude et promptitude du rapportage, chacune avec son dénominateur. Onze formations sanitaires sur trente-huit ont rapporté en août, et tout taux de couverture de ce mois est une affirmation sur ces onze-là.*

## 2.1 Deux indicateurs, et aucun n'est dans le rapport

Ouvrez presque n'importe quel rapport sanitaire de routine de ce secteur et vous y trouverez la couverture, la charge de cas, les taux de guérison et les ruptures de stock. Vous y trouverez très rarement les deux chiffres qui disent quelle part du district ces figures décrivent.

Ils sont bon marché. Les deux sortent de l'extraction que vous avez déjà — enfin, l'un des deux — et les deux changent la manière dont il faut lire tous les autres nombres du rapport.

## 2.2 La complétude du rapportage

Le taux de rapportage est la part des rapports attendus qui ont effectivement été transmis. Le mot qui travaille est **attendus**.

```
import pandas as pd

vax = pd.read_csv("vaccination-coverage-2024.v1.csv", parse_dates=["period"])
vax["reported"] = vax["report_submitted"] == True

print(f"overall reporting rate: {vax['reported'].mean():.1%}")

by_month = vax.groupby(vax["period"].dt.strftime("%Y-%m"))["reported"].mean()
print((by_month * 100).round(0))
```

```
library(dplyr)
library(readr)

vax <- read_csv("vaccination-coverage-2024.v1.csv")

mean(vax$report_submitted)

vax |>
  mutate(month = format(period, "%Y-%m")) |>
  summarise(reporting_rate = mean(report_submitted), .by = month) |>
  arrange(month)
```

| Mois                 | Taux de rapportage |
|----------------------|--------------------|
| Janvier              | 92 %               |
| Février              | 82 %               |
| Mars                 | 74 %               |
| Avril à juillet      | 82 à 87 %          |
| <b>Août</b>          | <b>29 %</b>        |
| <b>Septembre</b>     | <b>45 %</b>        |
| Octobre              | 95 %               |
| Novembre et décembre | 76 à 87 %          |

Août et septembre ne sont pas une oscillation. Onze formations sanitaires sur trente-huit ont rapporté en août et dix-sept en septembre. **Tout taux de couverture d’août est une affirmation sur onze formations**, et le publier à côté de celui de juillet sans le dire revient à comparer deux districts différents.

Remarquez aussi que l’effondrement et la reprise sont simultanés dans tout le district. C’est la signature d’un événement de système — un stock de formulaires papier, une grève, un serveur, un superviseur parti — et non de formations qui défaillent individuellement. Une évaluation qui soulève 27 constats séparés pour août a mal lu le motif.

## 2.3 Décomposez toujours de trois façons

Un nombre unique masque tout ce qui mérite d’être su. Décomposez le taux de rapportage par période, par unité de rapportage et par type d’unité.

```
by_facility = vax.groupby("facility_id")["reported"].mean().sort_values()
print(by_facility.head(6))
print(f"{{(by_facility == 1).sum()}} facilities reported every month")

by_type = vax.groupby("facility_type")["reported"].mean()
print((by_type * 100).round(1))
```

```
vax |> summarise(rate = mean(report_submitted), .by = facility_id) |> arrange(rate) |>
  head(6)
```

```
vax |> summarise(rate = mean(report_submitted), .by = facility_type)
```

| Type de formation   | Taux de rapportage |
|---------------------|--------------------|
| Centre de santé     | 82,8 %             |
| Hôpital de district | 75,0 %             |
| Poste de santé      | 71,7 %             |

Les six pires formations sont toutes à 58 % — sept mois sur douze — et seules deux sur trente-huit ont rapporté tous les mois. Le gradient par type est le constat actionnable : **ce sont les postes de santé qui rapportent le moins bien**, ce qui désigne la distance de supervision, les effectifs ou le transport plutôt qu’une formation en particulier.

## 2.4 Le piège du dénominateur

L’erreur unique qui rend un taux de rapportage inutile.

```
# WRONG: the share of reports that were submitted, among reports that were submitted
wrong = vax[vax["reported"]]["reported"].mean() # always 1.0
```

```
# WRONG
vax |> filter(report_submitted) |> summarise(rate = mean(report_submitted))
```

Évident une fois écrit. Cela ne l'est pas quand le filtre est six lignes plus haut dans une chaîne, et c'est exactement ce qui arrive lorsqu'on lit une extraction DHIS2 ne contenant que les rapports transmis. **Si l'extraction ne comporte aucune ligne pour les unités non déclarantes, le taux de rapportage ne peut pas y être calculé du tout** — il vous faut la liste des unités attendues, prise ailleurs.

```
EXPECTED_FACILITIES = pd.read_csv("facility-master-list.csv")["facility_id"]

expected = len(EXPECTED_FACILITIES) * vax["period"].nunique()
submitted = vax.loc[vax["reported"], ["facility_id", "period"]].drop_duplicates().shape[0]
print(f"{submitted} of {expected} expected facility-months")

expected <- nrow(facility_master) * n_distinct(vax$period)
submitted <- vax |> filter(report_submitted) |> distinct(facility_id, period) |> nrow()
```

La liste de référence fait autorité, exactement comme la base administrative de population dans le cours précédent. Une formation fermée et une formation qui a cessé de rapporter sont identiques dans l'extraction et constituent des constats totalement différents, et seule la liste de référence les distingue.

## 2.5 La promptitude exige un champ que vous n'avez peut-être pas

La promptitude est la part des rapports transmis avant l'échéance, et le retard médian des autres.

```
submissions["days_late"] = (
    submissions["submitted_on"] - submissions["deadline"]
).dt.days

print(f"on time: {(submissions['days_late'] <= 0).mean():.1%}")
print(f"median days late, of those late: "
      f"{submissions.loc[submissions['days_late'] > 0, 'days_late'].median():.0f}")

submissions <- submissions |>
  mutate(days_late = as.integer(submitted_on - deadline))

c(on_time = mean(submissions$days_late <= 0),
  median_late = median(submissions$days_late[submissions$days_late > 0]))
```

**Cette extraction ne permet pas ce calcul**, et le dire est en soi un constat d'évaluation. Il n'y a pas d'horodatage de transmission dans le fichier — seulement `report_submitted`, un indicateur sans date. DHIS2 enregistre un `lastUpdated` par ensemble de valeurs et une date de complétude par enregistrement, si bien que le champ existe dans le système et n'a pas été repris dans l'export.

Consignez-le tel quel.

**Constat.** La promptitude ne peut pas être évaluée. L'extraction mensuelle porte un indicateur de transmission mais aucune date de transmission, alors que DHIS2 en enre-

gistre une. **Action corrective.** Ajouter la date de complétude à l'extraction standard. **Responsable.** Point focal SNIS. **Échéance.** Prochaine extraction trimestrielle.

Une dimension que vous ne pouvez pas mesurer est un constat sur le système de rapportage, non une lacune de votre évaluation. Rapportez-la dans le même tableau que le reste.

## 2.6 Lisez tous les autres chiffres à travers ces deux-là

L'intérêt de les calculer en premier tient à ce qu'ils vous autorisent à dire ensuite.

- **La couverture, à 76,5 % de rapportage.** Un total de district est une borne inférieure, pas une estimation. Écrivez « parmi les formations ayant rapporté » dans le libellé, à chaque fois.
- **Une tendance mensuelle.** La baisse d'août est une baisse de rapportage jusqu'à preuve du contraire. Tracez le taux de rapportage sur le même graphique que l'indicateur, et la question se répond d'elle-même.
- **Un classement de formations.** Des formations ayant rapporté sept mois sur douze ne peuvent pas être classées face à des formations en ayant rapporté douze. Soit vous limitez le classement à un ensemble commun de mois, soit vous classez sur une moyenne mensuelle en affichant le décompte.

```
monthly = (
    vax[vax["reported"]]
    .groupby(vax["period"].dt.strftime("%Y-%m"))
    .apply(lambda g: g["doses_administered"].sum() / g["target_population"].sum())
    .rename("coverage")
    .to_frame()
)
monthly["reporting_rate"] = by_month
print(monthly.round(3))
```

```
vax |>
  mutate(month = format(period, "%Y-%m")) |>
  summarise(
    coverage = sum(doses_administered[report_submitted]) /
               sum(target_population[report_submitted]),
    reporting_rate = mean(report_submitted),
    .by = month
  )
```

Deux colonnes, toujours adjacentes. Un taux de couverture sans son taux de rapportage est un nombre sans étiquette, et l'étiquette est ce qui l'empêche d'être mal lu.

## 2.7 La suite

Complétude et promptitude disent si un rapport est arrivé. La leçon suivante dit s'il était juste — le recomptage face au registre source, le facteur de vérification qu'il produit, et la bande de tolérance qui en fait une décision.

Deuxième partie

**Confronter à la source**

## CHAPITRE 3

# Le recomptage, et le nombre qu'il produit

Leçon 3 sur 8 · 75 min

*Un facteur de vérification est le recomptage divisé par le chiffre rapporté. Le district est à 1,012, ce que personne ne contesterait, et une école sur vingt-quatre est à 1,42.*

### 3.1 L'exactitude est la dimension qui exige un déplacement

Les quatre autres dimensions sortent de l'extraction. Celle-ci non. L'exactitude signifie « le chiffre rapporté correspond-il à la source », et la source est un registre papier dans une armoire, une fiche de pointage, ou une tablette qui n'a plus synchronisé depuis mars.

La méthode est la plus ancienne de ce cours et personne n'a fait mieux — **allez le recompter**. Puis divisez.

### 3.2 Le facteur de vérification

```
verification factor = recount / reported
```

Ce que vous avez compté à la source, divisé par ce qui a été remonté. Un facteur de 1,00 signifie que le rapport correspondait au registre. C'est un rapport, non un pourcentage d'erreur, et le mettre dans le bon sens compte au moment de l'écrire.

**Les deux sens sont des constats, et ce ne sont pas les mêmes.**

- **Facteur supérieur à 1** — le registre contient plus que ce qui a été rapporté. Sous-déclaration. Généralement une perte de transcription ou d'agrégation : une colonne non additionnée, une page sautée, une extraction électronique qui a laissé tomber des valeurs qu'elle ne reconnaissait pas. Le programme a fait plus de travail qu'on ne lui en attribue.
- **Facteur inférieur à 1** — le rapport revendique plus que le registre ne soutient. Surdéclaration. Parfois un double comptage ou un écart de définition ; occasionnellement quelque chose de plus grave. C'est le sens qui fait suspendre une subvention, donc celui qu'il faut affirmer avec le plus de précaution.

### 3.3 Calculez-le sur un registre dont vous disposez

Prenez le registre de présence scolaire. L'extraction électronique porte des marques `true` et `false` ; une école a aussi enregistré certains jours en Y et N, que la conversion booléenne de l'extraction n'a pas reconnus. Une visite de vérification lisant le registre papier compterait ces jours comme des présences. L'extraction ne l'a pas fait.

Cela donne un recomptage réel et un chiffre rapporté réel, sur les mêmes données.

```
import pandas as pd

attendance = pd.read_csv("school-attendance-2024.v1.csv")
roster = pd.read_csv("school-roster-2024.v1.csv")

# The two students on the roster twice are ambiguous; exclude them from the
```

```
# verification and say so, rather than assigning them arbitrarily.
duplicated = roster["student_id"].duplicated(keep=False)
clean_roster = roster[~duplicated]

marks = attendance.merge(
    clean_roster[["student_id", "school_id"]], on="student_id",
    how="inner", validate="many_to_one",
)

vf = (
    marks.assign(
        reported=marks["present"] == "true",
        recount=marks["present"].isin(["true", "Y"]),
    )
    .groupby("school_id")[["recount", "reported"]]
    .sum()
)
vf["verification_factor"] = vf["recount"] / vf["reported"]
print(vf.sort_values("verification_factor", ascending=False).head())

library(dplyr)

clean_roster <- roster |>
  group_by(student_id) |> filter(n() == 1) |> ungroup()

vf <- attendance |>
  inner_join(select(clean_roster, student_id, school_id), by = "student_id",
    relationship = "many-to-one") |>
  summarise(
    recount = sum(present %in% c("true", "Y")),
    reported = sum(present == "true"),
    .by = school_id
  ) |>
  mutate(verification_factor = recount / reported) |>
  arrange(desc(verification_factor))
```

| École           | Recomptage    | Rapporté      | Facteur      |
|-----------------|---------------|---------------|--------------|
| SCH09           | 2 456         | 1 734         | <b>1,416</b> |
| SCH01           | 2 945         | 2 945         | 1,000        |
| SCH02           | 2 569         | 2 569         | 1,000        |
| ... 22 autres   |               |               | 1,000        |
| <b>District</b> | <b>61 764</b> | <b>61 042</b> | <b>1,012</b> |

### 3.4 Le chiffre du district est le piège

Lisez la dernière ligne d'abord, car c'est celle qui figure dans la plupart des rapports. **Le facteur de vérification du district vaut 1,012** — un peu plus d'un pour cent — et aucun auditeur au monde ne soulèverait de constat là-dessus.

Lisez maintenant la première ligne. Une école sur vingt-quatre a sous-déclaré sa présence de 42 %, et cela est invisible dans le total du district parce que les vingt-trois autres sont exactes.

C'est l'argument que le cours précédent avançait à propos de la moyenne des facteurs de vérification, et il mérite d'être avancé deux fois car c'est la manière la plus courante dont une évaluation délivre un satisfecit à un district qui a un vrai problème. **Rapportez la**

**distribution et les exceptions. La moyenne est la statistique la moins informative disponible.**

```
print(vf["verification_factor"].describe())
outliers = vf[(vf["verification_factor"] - 1).abs() > 0.05]
print(f"{len(outliers)} of {len(vf)} schools outside the tolerance band")
```

```
summary(vf$verification_factor)
vf |> filter(abs(verification_factor - 1) > 0.05)
```

### 3.5 La bande de tolérance

Une évaluation a besoin d'un seuil fixé avant de regarder, sans quoi chaque constat devient une négociation.

La convention retenue par la plupart des cadres bailleurs est **0,95 à 1,05** — à 5 % près c'est acceptable, au-delà c'est un constat. C'est une convention et non une loi, et deux ajustements sont légitimes.

- **Élargissez-la pour les petits nombres.** Une formation rapportant 12 cas a un facteur de 1,08 si un cas a été manqué. Sur de petits effectifs, employez aussi une différence absolue — « hors de 5 % *et* de plus de 5 unités ».
- **Resserrez-la quand il y a de l'argent.** Lorsque le chiffre déclenche un paiement — financement basé sur les résultats, remboursement par bénéficiaire — 5 % représentent beaucoup d'argent et la bande est en général plus étroite, fixée par le contrat et non par vous.

```
TOLERANCE = 0.05
MIN_ABSOLUTE = 5

vf["difference"] = vf["recount"] - vf["reported"]
vf["finding"] = (
    ((vf["verification_factor"] - 1).abs() > TOLERANCE)
    & (vf["difference"].abs() > MIN_ABSOLUTE)
)
print(vf[vf["finding"]])
```

```
vf <- vf |>
  mutate(
    difference = recount - reported,
    finding = abs(verification_factor - 1) > 0.05 & abs(difference) > 5
  )
```

**Inscrivez la bande dans le protocole avant la visite.** Un seuil choisi après avoir vu les chiffres n'est pas un seuil, et tout le monde dans la salle le saura.

### 3.6 Le recomptage est un protocole, pas une après-midi

L'arithmétique est triviale. Tout ce qui rend le nombre digne de confiance se joue en amont.

- **Fixez la période exactement.** « Mars » n'est pas une période. « Du 1er au 31 mars, par date de service et non par date de saisie » en est une.

- **Fixez la définition exactement.** Comptez ce que dit le numérateur de l'indicateur, non ce que dit l'en-tête de colonne du registre. La plupart des facteurs éloignés de 1 sont des écarts de définition, et ce ne sont les erreurs de personne.
- **Recomptez indépendamment.** La personne qui a compilé le rapport ne doit pas faire le recomptage. Non par défiance — parce qu'elle reproduira la même interprétation, et c'est l'interprétation que vous testez.
- **Notez à partir de quoi vous avez compté.** Registre, fiche de pointage, cartes individuelles. Deux sources d'une même formation se contredisent souvent, et celle que vous avez comptée fait partie du constat.

### 3.7 Ce qu'un facteur de vérification ne peut pas dire

Trois limites à énoncer dans le rapport, sinon quelqu'un sur-interprétera le nombre.

**Il ne dit pas que le registre est juste.** Un facteur de 1,00 signifie que le rapport correspond au registre. Si le registre a été rempli en fin de semaine de mémoire, les deux sont faux ensemble et le facteur reste muet.

**Il ne se généralise pas au-delà des formations visitées,** sauf si vous avez échantillonné correctement. C'est la leçon suivante.

**Ce n'est pas une accusation.** Un facteur de 1,42 dans une école, avec vingt-trois autres à exactement 1,00, a la silhouette d'un défaut de système — la convention de saisie d'une école, non reconnue par l'extraction — et non celle d'un directeur gonflant ses chiffres. L'écrire de la seconde façon garantit que le tour suivant sera pire.

### 3.8 La suite

Vous ne pouvez pas visiter trente-huit formations. La leçon suivante montre comment choisir celles que vous visiterez, combien suffisent, et ce qu'un échantillon de six vous autorise à dire des trente-deux autres.

## CHAPITRE 4

# Choisir les six formations sanitaires à visiter

Leçon 4 sur 8 · 75 min

*Trois stratégies d'échantillonnage qui répondent à trois questions différentes, combien d'unités suffisent, et la phrase que vous avez le droit d'écrire sur les trente-deux formations non visitées.*

### 4.1 Vous avez une semaine et trente-huit formations sanitaires

Une visite de vérification coûte un véhicule, une journée, deux personnes et du carburant. Le budget en couvre six. Tout dans cette leçon découle de cette arithmétique, et le premier point à clarifier est que **le choix des six détermine ce que votre constat aura le droit de signifier.**

Trois stratégies, trois questions, et les confondre est la défaillance méthodologique la plus courante d'une évaluation.

| Stratégie                    | Répond à                                               | Ne peut pas répondre à         |
|------------------------------|--------------------------------------------------------|--------------------------------|
| Aléatoire                    | « Quelle est la qualité des données du district ? »    | « Où sont les problèmes ? »    |
| Fondée sur le risque         | « Les formations suspectées le sont-elles vraiment ? » | « Quelle est la qualité du dis |
| Recensement des plus grosses | « Le total du district est-il juste ? »                | Quoi que ce soit sur les petit |

### 4.2 Aléatoire — la seule qui se généralise

Si vous voulez une affirmation sur le district, l'échantillon doit être tiré sans référence à ce que vous vous attendez à trouver.

```
import pandas as pd

facilities = vax[["facility_id", "facility_type"]].drop_duplicates()

sample = facilities.sample(n=6, random_state=20260728)
print(sample)

set.seed(20260728)

sample <- facilities |> dplyr::slice_sample(n = 6)
```

**Fixez la graine et consignez-la.** Une évaluation dont l'échantillon n'est pas reproductible appelle la question de savoir si les formations ont été choisies après que quelqu'un a su ce qu'elles contenaient, et la graine y répond en une ligne.

Stratifiez lorsqu'un sous-groupe compte. Les postes de santé rapportent le moins bien, si bien qu'un simple tirage aléatoire de six pourrait n'en contenir aucun.

```
sample = (
    facilities.groupby("facility_type", group_keys=False)
    .apply(lambda g: g.sample(n=min(2, len(g)), random_state=20260728))
)
```

```
sample <- facilities |>
  group_by(facility_type) |>
  slice_sample(n = 2) |>
  ungroup()
```

Stratifier par type donne une affirmation sur chaque type en plus du district, et cela ne coûte rien.

### 4.3 Fondée sur le risque — là où sont les problèmes

Si le but est de corriger plutôt que de décrire, choisissez les formations que l'analyse de bureau a déjà signalées. Les deux leçons suivantes sont entièrement consacrées à produire ce classement.

```
risk = (
  vax.groupby("facility_id")
  .agg(reporting_rate=("reported", "mean"))
  .assign(round_share=lambda d: d.index.map(round_number_share))
)
risk["score"] = (1 - risk["reporting_rate"]) + risk["round_share"]
print(risk.sort_values("score", ascending=False).head(6))
```

```
risk <- vax |>
  summarise(reporting_rate = mean(report_submitted), .by = facility_id) |>
  left_join(round_shares, by = "facility_id") |>
  mutate(score = (1 - reporting_rate) + round_share) |>
  arrange(desc(score))
```

Cette approche trouve plus de problèmes par journée-véhicule que le tirage aléatoire, raison pour laquelle la plupart des évaluations réelles l'emploient. Elle **ne se généralise pas**, et le rapport doit le dire.

Six formations ont été retenues sur la base des signalements de l'analyse de bureau. Les constats les décrivent et ne sont pas représentatifs du district.

Tout rapport d'évaluation qui omet cette phrase puis cite un facteur de vérification au niveau du district formule une affirmation que son protocole ne soutient pas.

### 4.4 Recensement des plus grosses

Une troisième option, souvent la plus utile pour un total. Ordonnez les formations par volume et visitez-en assez pour couvrir l'essentiel du numérateur.

```
volume = (
  vax[vax["reported"]]
  .groupby("facility_id")["doses_administered"].sum()
  .sort_values(ascending=False)
)
share = volume.cumsum() / volume.sum()
print(share.head(10).round(3))
```

```
volume <- vax |>
  filter(report_submitted) |>
```

```
summarise(doses = sum(doses_administered), .by = facility_id) |>
  arrange(desc(doses)) |>
  mutate(cumulative = cumsum(doses) / sum(doses))
```

Si huit formations représentent la moitié des doses, vérifier ces huit borne l'erreur du total du district quoi que fassent les trente autres. **C'est le bon protocole quand la question est « puis-je me fier au total que je remonte au bailleur »** — et le mauvais quand la question porte sur la qualité du service dans les postes reculés, puisqu'il les ignore systématiquement.

## 4.5 Combien suffisent

Il n'y a pas de réponse unique, mais il y a une façon d'y penser qui résiste à la contestation.

Pour une **question par oui ou non** — le rapportage de cette formation est-il acceptable — vous estimez une proportion, et la précision s'améliore comme la racine carrée de l'échantillon. Six formations sur trente-huit donnent un intervalle de confiance si large que presque aucun résultat ne se distingue d'un autre. Soyez honnête là-dessus plutôt que de rapporter « 33 % des formations présentent des constats » à partir d'un échantillon de six.

Pour **détecter si un problème existe**, les petits échantillons sont bien plus capables. Si 30 % des formations ont un défaut, la probabilité que six formations tirées au hasard n'en contiennent aucune est d'environ 12 % — un échantillon de six sans constat est donc un indice raisonnable de l'absence d'un problème *répandu*, et aucun indice sur un problème rare.

```
p = 0.30
n = 6
print(f"chance of finding none: {(1 - p) ** n:.1%}")
```

```
(1 - 0.30)^6
```

Cette seule ligne mérite de figurer dans le rapport. Elle convertit « nous en avons visité six sans rien trouver » en une affirmation assortie d'un nombre.

## 4.6 LQAS — conçu exactement pour cela

L'échantillonnage par lots pour l'assurance qualité formalise le paragraphe précédent, et ce secteur l'emploie déjà pour les enquêtes de couverture. L'idée est de cesser de vouloir *estimer* et de commencer à vouloir *décider* — fixer un seuil, une taille d'échantillon et une règle de décision, puis accepter ou rejeter.

Visiter 12 formations. Si 3 ou plus ont un facteur de vérification hors de 0,95-1,05, le rapportage du district est classé comme nécessitant une action corrective.

L'avantage est que l'échantillon peut être petit parce que la question l'est. Le coût est que vous obtenez un verdict et non un chiffre, et que quelqu'un réclamera le chiffre malgré tout. Dites d'emblée pour lequel vous avez conçu.

## 4.7 Combinez, et étiquetez

Les évaluations réelles emploient souvent les trois, et c'est très bien tant que le rapport les sépare.

```
plan = pd.DataFrame({
    "facility_id": ["FAC012", "FAC033", "FAC004", "FAC019", "FAC007", "FAC021"],
    "reason": ["random", "random", "risk: reporting 58%", "risk: round numbers",
              "volume: top 5", "volume: top 5"],
})
```

```
plan <- tibble::tribble(
  ~facility_id, ~reason,
  "FAC012", "random",
  "FAC033", "random",
  "FAC004", "risk: reporting 58%",
  "FAC019", "risk: round numbers",
  "FAC007", "volume: top 5",
  "FAC021", "volume: top 5"
)
```

**La colonne reason est la section méthodologie.** Avec elle, un lecteur sait quels constats se généralisent et lesquels non. Sans elle, le rapport a un échantillon et trois affirmations incompatibles, et un relecteur qui le remarque les écartera toutes.

## 4.8 La suite

L'échantillonnage fondé sur le risque exige un classement de risque, et vous n'avez pour l'instant que des taux de rapportage. Les deux leçons suivantes construisent le reste à partir de l'extraction que vous détenez déjà — des tendances qui bougent de façon impossible, des valeurs qui se répètent, et des chiffres qui trahissent un nombre que personne n'a mesuré.

## Troisième partie

# Repérer les sites qui méritent une visite

## CHAPITRE 5

# Comparer une formation sanitaire à son propre passé

Leçon 5 sur 8 · 75 min

*Rapport à la médiane glissante, le problème des petits effectifs qui rend les plus grandes aberrations insignifiantes, et la série si stable qu'elle mérite un second regard.*

### 5.1 La cohérence est une comparaison, et vous en choisissez le terme

La dimension de cohérence demande si les nombres se tiennent. En pratique cela revient à comparer chaque valeur rapportée à quelque chose, et il y a deux candidats.

**À ses pairs.** La formation A a rapporté 40 doses ; la médiane du district est de 25. Tentant et généralement faux — les formations diffèrent énormément par la taille de leur bassin, et une grande formation sera signalée chaque mois pour cause de grandeur.

**À son propre passé.** La formation A a rapporté 40 doses ce mois-ci, contre sa propre médiane annuelle de 25. Voilà une comparaison qui vaut la peine, car une formation est un témoin raisonnable d'elle-même.

Employez la seconde. Tous les contrôles de cette leçon confrontent une formation à son propre historique.

### 5.2 Rapport à la médiane

La statistique utilise la plus simple, et elle résiste à la brièveté de ces séries.

```
import pandas as pd

reported = vax[vax["reported"]].copy()

reported["own_median"] = reported.groupby(["facility_id", "antigen"])["doses_administered"]
    .transform("median")
reported["ratio"] = reported["doses_administered"] / reported["own_median"]

print(reported.nlargest(5, "ratio")[
    ["facility_id", "antigen", "period", "doses_administered", "own_median", "ratio"]
])

library(dplyr)

reported <- vax |>
  filter(report_submitted) |>
  mutate(own_median = median(doses_administered),
         ratio = doses_administered / own_median,
         .by = c(facility_id, antigen))

reported |> slice_max(ratio, n = 5)
```

**La médiane, pas la moyenne.** La moyenne est tirée par l'aberration même que vous cherchez, si bien qu'une valeur peut gonfler sa propre base de comparaison et se dissimuler.

C'est la raison pour laquelle les contrôles de plausibilité SMART emploient des statistiques robustes.

Là où la série est assez longue, préférez une médiane *glissante*, pour qu'un véritable changement de niveau saisonnier ne signale pas tous les mois suivants.

```
reported = reported.sort_values(["facility_id", "antigen", "period"])
reported["rolling_median"] = (
    reported.groupby(["facility_id", "antigen"])["doses_administered"]
    .transform(lambda s: s.rolling(6, min_periods=3, center=True).median())
)

reported <- reported |>
  arrange(facility_id, antigen, period) |>
  mutate(rolling_median = zoo::rollapply(doses_administered, 6, median,
    partial = TRUE, align = "center"),
    .by = c(facility_id, antigen))
```

### 5.3 Le problème des petits effectifs

Exécutez le contrôle de rapport sur cette extraction et voici le haut de la liste.

| Formation | Antigène | Mois      | Valeur | Médiane propre | Rapport |
|-----------|----------|-----------|--------|----------------|---------|
| FAC013    | mcv2     | décembre  | 5      | 3              | 1,67    |
| FAC037    | opv3     | septembre | 13     | 8              | 1,62    |
| FAC030    | mcv2     | octobre   | 8      | 5              | 1,60    |

La plus grande aberration de tout le district est **une formation qui a administré cinq secondes doses de vaccin antirougeoleux au lieu de trois**. Deux doses. Personne n'ira enquêter là-dessus, et si votre rapport d'évaluation ouvre là-dessus, personne ne lira la suite.

C'est la défaillance caractéristique du signalement par rapport, et elle est garantie plutôt que malchanceuse — un rapport a un petit dénominateur en bas de la plage, donc les rapports les plus extrêmes proviennent toujours des plus petits effectifs. Le contrôle fait exactement ce qu'on lui a demandé et c'est la question qui était mauvaise.

### 5.4 Adjoignez un plancher absolu au rapport

```
RATIO_HIGH, RATIO_LOW, MIN_ABSOLUTE = 2.0, 0.5, 10

reported["difference"] = reported["doses_administered"] - reported["own_median"]
reported["flag"] = (
    ((reported["ratio"] > RATIO_HIGH) | (reported["ratio"] < RATIO_LOW))
    & (reported["difference"].abs() >= MIN_ABSOLUTE)
)
print(f"{reported['flag'].sum()} flags from {len(reported)} facility-antigen-months")

reported <- reported |>
  mutate(
    difference = doses_administered - own_median,
    flag = (ratio > 2 | ratio < 0.5) & abs(difference) >= 10
  )
```

Appliqué ici, rien du tout ne dépasse le double de sa propre médiane, et onze couples formation-antigène-mois tombent sous la moitié de la leur. Le mouvement réel le plus important est le penta3 de FAC005 en octobre — 13 doses contre une médiane de 43, une baisse réelle de trente.

**Voilà le constat, et il tient en une ligne.** Un contrôle qui produit un seul élément à investiguer sur 2 094 observations fonctionne; un contrôle qui en produit deux cents est un contrôle que personne n'exécutera deux fois.

## 5.5 Le sens n'est pas symétrique

Une baisse et un pic ne veulent pas dire la même chose, et une évaluation qui les traite pareillement manque l'essentiel de ce qui se passe.

- **Une baisse** est généralement réelle ou liée au rapportage. Rupture de stock, absence de personnel, grève, formation qui s'est mise à rapporter à un autre système. Interrogez le mois, pas le nombre.
- **Un pic** est généralement une campagne, une sortie de proximité ou un rattrapage après rupture — toutes légitimes — ou un double comptage, qui ne l'est pas. Vérifiez si les mois encadrants baissent en conséquence, signature d'un problème de bornes de période plutôt que d'une activité supplémentaire.

```
reported["previous"] = reported.groupby(["facility_id", "antigen"])[
    "doses_administered"].shift(1)
reported["next"] = reported.groupby(["facility_id", "antigen"])[
    "doses_administered"].shift(-1)

spike_and_dip = reported[
    (reported["ratio"] > 1.5)
    & (reported["previous"] < reported["own_median"] * 0.7)
]
```

```
reported <- reported |>
  mutate(previous = lag(doses_administered),
         next_value = lead(doses_administered),
         .by = c(facility_id, antigen))
```

Un mois élevé précédé d'un mois faible est très souvent le travail d'un mois consigné dans la remontée du suivant. C'est un constat d'actualité déguisé en constat de cohérence, et le remède est une échéance, pas un recomptage.

## 5.6 Des séries trop stables

Le contrôle inverse, celui qu'on oublie. La prestation de service réelle est bruitée. Une formation rapportant un chiffre presque identique chaque mois mérite un coup d'œil.

```
stability = (
    reported.groupby(["facility_id", "antigen"])["doses_administered"]
    .agg(["mean", "std", "size"])
)
stability["cv"] = stability["std"] / stability["mean"]
print(stability[stability["size"] >= 8].nsmallest(5, "cv"))
```

```
reported |>
  summarise(mean = mean(doses_administered),
            cv = sd(doses_administered) / mean(doses_administered),
            n = n(),
            .by = c(facility_id, antigen)) |>
  filter(n >= 8) |>
  slice_min(cv, n = 5)
```

La série la plus stable de cette extraction a un coefficient de variation d'environ 0,06 — six pour cent d'un mois à l'autre. C'est faible, et ce **n'est la preuve de rien**. Une grande formation à bassin stable et à jour de consultation régulier peut aisément produire cela.

Énoncez la limite franchement, car ce contrôle est le plus facilement détourné : un faible coefficient de variation met une formation sur la liste des visites. Il ne figure pas au rapport comme un constat, et surtout pas dans une phrase contenant le mot « fabriqué ».

## 5.7 Classez les formations, ne signalez pas des lignes

Le produit de cette leçon n'est pas une liste de mois signalés. C'est un classement d'unités de rapportage, ce dont la leçon d'échantillonnage avait besoin.

```
risk = (
  reported.groupby("facility_id")
  .agg(flags=("flag", "sum"), months=("flag", "size"))
  .assign(flag_rate=lambda d: d["flags"] / d["months"])
  .sort_values("flag_rate", ascending=False)
)
print(risk.head(6))
```

```
risk <- reported |>
  summarise(flags = sum(flag), months = n(), .by = facility_id) |>
  mutate(flag_rate = flags / months) |>
  arrange(desc(flag_rate))
```

Une formation avec quatre signalements sur douze mois est candidate à une visite. Un mois signalé isolé dans une formation par ailleurs régulière est une question pour un appel téléphonique. **L'unité d'action est la formation, parce que c'est là que va le véhicule.**

## 5.8 La suite

Les contrôles de tendance comparent un nombre à son propre passé. La leçon suivante regarde à l'intérieur du nombre lui-même — le dernier chiffre, l'attraction des multiples de cinq, et les limites honnêtes de ce que cela peut dire de la façon dont une mesure a été produite.

## CHAPITRE 6

# Ce que dit le dernier chiffre

Leçon 6 sur 8 · 75 min

*Préférence de chiffres, attraction des valeurs rondes, et l'étalonnage qui sépare un constat du bruit. Les tailles d'une équipe se terminent par ,0 ou ,5 dans soixante-neuf pour cent des cas ; la formation la plus suspecte du district se révèle être du hasard.*

### 6.1 Les mesures ont une texture

Un nombre produit par la lecture d'un instrument a une texture particulière — son dernier chiffre est quasi uniforme, car la valeur vraie a autant de chances de tomber n'importe où dans la plus petite division de l'échelle.

Un nombre produit par estimation, arrondi ou souvenir a une autre texture. Il se regroupe sur les cinq et les zéros, sur les années entières, sur les multiples de dix. Ce regroupement est la **préférence de chiffres**, et il se mesure sans se déplacer.

C'est l'outil principal de la dimension d'intégrité, et cette leçon consacre autant de temps à ce qu'il ne peut pas montrer qu'à ce qu'il montre.

### 6.2 Le contrôle

```
import pandas as pd

smart = pd.read_csv("smart-nutrition-survey-2024.v1.csv")

smart["terminal"] = (smart["height_cm"] * 10).round() % 10
by_team = (
    smart.assign(rounded=smart["terminal"].isin([0, 5]))
    .groupby("team")
    .agg(rounded_share=("rounded", "mean"), n=("rounded", "size"))
)
print((by_team["rounded_share"] * 100).round(0))
```

```
library(dplyr)

smart |>
  mutate(terminal = round(height_cm * 10) %% 10,
         rounded = terminal %in% c(0, 5)) |>
  summarise(rounded_share = mean(rounded), n = n(), .by = team)
```

| Équipe | Tailles finissant par ,0 ou ,5 | Enfants mesurés |
|--------|--------------------------------|-----------------|
| 1      | 18 %                           | 217             |
| 2      | 69 %                           | 248             |
| 3      | 18 %                           | 248             |
| 4      | 23 %                           | 217             |

Deux chiffres terminaux possibles sur dix font 20 % en l'absence de préférence. Les équipes 1, 3 et 4 sont à 18 %, 18 % et 23 % — ce à quoi ressemble une mesure. L'équipe 2 est à 69 %.

Ce n'est pas un résultat marginal et cela n'exige aucun test. Rien dans une population ne fait tomber les tailles des enfants d'une équipe sur des demi-centimètres trois fois et demie plus souvent que celles de tout le monde. C'est la façon dont l'équipe a lu la toise — arrondir au demi-centimètre le plus proche au lieu de lire le millimètre.

### 6.3 L'attraction des âges, et pourquoi ce n'est la faute de personne

Le même phénomène sur les âges, où il est quasi universel dans ce secteur.

```
ages = smart["age_months"].dropna()
whole_years = (ages % 12 == 0).mean()

print(f"{whole_years:.0%} of ages fall on an exact multiple of 12 months")
print(ages.value_counts().reindex([22, 23, 24, 25, 26]))
```

```
smart |>
  filter(!is.na(age_months)) |>
  summarise(whole_years = mean(age_months %% 12 == 0))

smart |> count(age_months) |> filter(age_months %in% 22:26)
```

| Âge en mois | 22 | 23 | 24 | 25 | 26 |
|-------------|----|----|----|----|----|
| Enfants     | 21 | 15 | 66 | 19 | 17 |

Soixante-six enfants à exactement vingt-quatre mois contre quinze et dix-neuf de part et d'autre, et quatre-vingts à trente-six contre vingt-huit et dix-sept. Sur l'ensemble de l'enquête, **24 % des âges tombent sur une année entière exacte, contre environ 9 % attendus** si les âges étaient répartis uniformément.

Cela compte pour une raison précise et concrète, non par curiosité. Les normes de croissance de l'OMS changent de table de référence à 24 mois, et les groupes d'âge SMART sont bornés aux multiples de 12 mois — l'attraction dépose donc un gros paquet d'enfants exactement sur les bornes où la classification change.

Et la cause n'est presque jamais la négligence. Les actes de naissance ne sont pas universels ; une accompagnante à qui l'on demande l'âge d'un enfant répondra « deux ans » ; un calendrier d'événements locaux aide mais a ses limites. **Consignez-le comme un constat sur la méthode de détermination de l'âge**, avec une action corrective portant sur les calendriers et le questionnement, non comme une erreur d'enquêteur.

### 6.4 Étalonnez sur ce que produit le hasard

Vient maintenant ce qui sépare un contrôle exploitable d'une machine à accuser. Exécutez le contrôle des nombres ronds sur l'extraction vaccinale.

```
reported = vax[vax["reported"]].copy()
base_rate = (reported["doses_administered"] % 10 == 0).mean()

by_facility = (
  reported.assign(round_number=reported["doses_administered"] % 10 == 0)
  .groupby("facility_id")
  .agg(rounds=("round_number", "sum"), n=("round_number", "size"))
)
by_facility["share"] = by_facility["rounds"] / by_facility["n"]
```

```
print(f"district base rate: {base_rate:.3f}")
print(by_facility.nlargest(3, "share"))

base_rate <- mean(vax$doses_administered[vax$report_submitted] %% 10 == 0)

by_facility <- vax |>
  filter(report_submitted) |>
  summarise(rounds = sum(doses_administered %% 10 == 0), n = n(), .by = facility_id) |>
  mutate(share = rounds / n) |>
  arrange(desc(share))
```

Le taux de base du district est de 10,4 % — exactement l'attendu. La pire formation, FAC020, rapporte un nombre rond 14 fois sur 60, soit 23 %. Plus du double du taux du district. En soi cela ressemble à un constat.

Testez-le.

```
from scipy.stats import binomtest

worst = by_facility.nlargest(1, "share").iloc[0]
p = binomtest(int(worst["rounds"]), int(worst["n"]), base_rate,
              alternative="greater").pvalue
print(f"p = {p:.4f}, Bonferroni across 38 facilities: {min(1, p * 38):.3f}")

worst <- by_facility[1, ]
p <- binom.test(worst$rounds, worst$n, base_rate, alternative = "greater")$p.value
c(p = p, bonferroni = min(1, p * 38))
```

**p = 0,003 pris isolément, et 0,11 après correction pour avoir regardé trente-huit formations.** Trois formations passent sous  $p = 0,05$  sans correction, et 1,9 sont attendues par hasard. Trois contre 1,9, cela ne veut rien dire.

Une seconde raison de laisser tomber. Les doses mensuelles de FAC020 vont de cinq à treize, si bien que « se termine par zéro » signifie presque toujours que la valeur était exactement dix. Sur des effectifs aussi petits, le contrôle n'a aucune puissance pour distinguer l'arrondi de l'arithmétique.

Le résultat est : **aucun indice de préférence de chiffres dans le rapportage vaccinal**. C'est un vrai constat, il a pris quatre lignes, et c'est celui qu'il faut espérer.

## 6.5 Pourquoi le cas négatif compte davantage

Ce qui rend ce contrôle dangereux tient surtout à ce qu'on l'exécute presque toujours sur des données où rien ne cloche, avec l'espoir de trouver quelque chose.

Trois règles qui le maintiennent honnête.

- **Calculez le taux de base sur les données**, pas sur la théorie. Les chiffres terminaux ne sont pas uniformes quand les effectifs sont petits, quand un formulaire a une valeur par défaut, ou quand les doses viennent en flacons de dix.
- **Corrigez du nombre d'unités examinées**. Tester trente-huit formations à 5 % produit environ deux positifs sur des données propres, à chaque fois.
- **Exigez un mécanisme**. Les 69 % de l'équipe 2 en ont un — l'équipe a lu la toise au demi-centimètre le plus proche. Si vous ne savez pas nommer de mécanisme plausible, vous

avez un nombre et non un constat.

## 6.6 Ce que vous avez le droit d'écrire

La formulation n'est pas une politesse ; elle décide de la suite.

| N'écrivez pas                                     | Écrivez                                                                             |
|---------------------------------------------------|-------------------------------------------------------------------------------------|
| « Les données de FAC020 semblent fabriquées »     | « FAC020 rapporte des nombres ronds plus souvent que la moyenne »                   |
| « Les mesures de l'équipe 2 ne sont pas fiables » | « Les tailles de l'équipe 2 se terminent par ,0 ou ,5 dans 69 % des cas »           |
| « Les âges sont inexacts »                        | « 24 % des âges tombent sur une année entière exacte contre 9 % d'années entières » |

La colonne de droite dit ce qui a été observé, son ampleur, et le mécanisme avec lequel c'est compatible. Celle de gauche dit ce que quelqu'un a fait. Une seule des deux produit une action corrective, et une seule survit à sa présentation à l'équipe qu'elle décrit.

## 6.7 La suite

Vous avez maintenant des constats — un effondrement de complétude, un facteur de vérification de 1,42, une équipe qui arrondit ses tailles. Aucun n'est encore une cause, et aucun n'est actionnable. La leçon suivante montre comment passer d'un défaut à ce qui l'a produit, et pourquoi « il faut former le personnel » est presque toujours la mauvaise réponse.

## Quatrième partie

# Transformer les constats en changement

## CHAPITRE 7

# Du défaut à ce qui l'a produit

Leçon 7 sur 8 · 60 min

*Cinq catégories de cause, une règle d'arrêt pour les cinq pourquoi, et pourquoi « il faut former le personnel » est la réponse que l'on écrit quand personne n'a cherché. Testez la cause sur les données avant de l'inscrire au rapport.*

### 7.1 Un constat n'est pas une cause, et seule une cause se corrige

À ce stade vous avez des constats. Le rapportage s'est effondré à 29 % en août. Une école a sous-déclaré sa présence de 42 %. Une équipe d'enquête arrondit les tailles au demi-centimètre. Vingt-quatre pour cent des âges sont des années entières.

Chacun est un *quoi*. Aucun n'est un *pourquoi*, et un plan d'action corrective rédigé face à un *quoi* produit la phrase qui figure dans la plupart des rapports d'évaluation de ce secteur.

**Recommandation.** Former le personnel à la qualité des données.

Cette phrase s'écrit quand personne n'a cherché de cause. Elle est infalsifiable, elle est coûteuse, et elle ne changera rien au taux de rapportage d'août, parce que rien dans le mois d'août n'a été causé par un personnel ignorant son métier.

### 7.2 Cinq catégories, et elles n'exigent pas le même argent

Presque toute cause de qualité des données dans les données de programme de routine relève de l'une de cinq catégories. Nommer celle où vous êtes désigne qui doit corriger.

| Catégorie  | Ressemble à                                                                                           |
|------------|-------------------------------------------------------------------------------------------------------|
| Définition | Deux sites comptant des choses différentes sous un même nom d'indicateur                              |
| Outil      | Un formulaire incapable d'exprimer ce qui s'est passé, ou un système qui perd des valeurs             |
| Capacité   | Le bon outil mal utilisé, systématiquement, par quelqu'un à qui on ne l'a jamais montré               |
| Incitation | Un rapportage qui s'améliore quand on le regarde, ou des chiffres qui atteignent exactement une cible |
| Système    | Tout le monde échoue en même temps — ni formulaires, ni carburant, ni réseau, ni superviseur          |

Deux remarques sur ce tableau, et les deux reviennent à chaque évaluation.

**La définition est la plus grande catégorie et la moins soupçonnée.** Un facteur de vérification éloigné de 1 traduit plus souvent deux personnes comptant des choses différentes que quiconque comptant mal. C'est aussi la moins chère à corriger, et le remède — écrire la définition — prévient définitivement la récidive.

**La capacité est la plus petite catégorie et la plus fréquemment accusée,** car c'est la seule dont le remède est un atelier, et un atelier, cela s'achète.

### 7.3 Laissez les données réduire le champ avant d'interroger quiconque

La forme d'un défaut dit dans quelle catégorie il se trouve, et vous avez déjà les données pour la voir.

```
by_month = vax.groupby(vax["period"].dt.strftime("%Y-%m"))["reported"].mean()
by_facility = vax.groupby("facility_id")["reported"].mean()

print("spread across months: ", round(by_month.max() - by_month.min(), 2))
print("spread across facilities:", round(by_facility.max() - by_facility.min(), 2))

by_month <- vax |> mutate(m = format(period, "%Y-%m")) |>
  summarise(r = mean(report_submitted), .by = m)
by_facility <- vax |> summarise(r = mean(report_submitted), .by = facility_id)

c(months = diff(range(by_month$r)), facilities = diff(range(by_facility$r)))
```

Trois formes, trois catégories.

- **Concentrée dans le temps, étalée sur les unités.** Août, partout. C'est une cause système — quelque chose a lâché pour tout le district d'un coup. Vingt-sept constats par formation seraient vingt-sept réponses fausses.
- **Concentrée sur une unité, étalée dans le temps.** Le codage Y/N de SCH09, tous les mois. C'est une cause d'outil ou de définition locale à un site.
- **Étalée sur les deux.** L'attraction des âges, toutes équipes, toute l'année. C'est une cause de méthode, inhérente à la façon même dont la mesure est prise.

**Écrivez la forme avant de proposer une cause.** Cela élimine gratuitement trois des cinq catégories, et c'est l'étape qui empêche une défaillance système d'être consignée comme vingt-sept défaillances de capacité.

## 7.4 Puis interrogez le formulaire

Pour tout ce qui est local, regardez l'instrument avant de regarder la personne.

- Quelles valeurs le formulaire peut-il réellement accepter ? Les registres de SCH09 offraient des cases Y/N ; la conversion booléenne de l'extraction reconnaissait `true/false`. **Ni l'enseignant ni l'analyste n'a mal fait**, et aucune formation n'y remédie.
- Le champ est-il obligatoire ? Un champ facultatif avec 40 % de non-réponse est une décision de conception de formulaire, pas un problème de conformité.
- Que dit l'intitulé du champ ? Une colonne intitulée « cas » recueille ce que la personne qui la remplit croit être un cas.

## 7.5 Les cinq pourquoi, avec une règle d'arrêt

La technique est classique. Ce qui manque d'ordinaire est une règle d'arrêt, et c'est pourquoi les cinq pourquoi se terminent si souvent sur « le personnel n'est pas motivé ».

**Arrêtez-vous quand vous atteignez quelque chose qu'une personne nommée peut changer avec un budget que vous pouvez estimer.** Au-delà, c'est de la philosophie.

La présence a été sous-déclarée de 42 % à SCH09. — *Pourquoi ?* L'extraction a lu Y comme une valeur manquante. — *Pourquoi ?* Le registre employait Y/N et le formulaire attendait `true/false`. — *Pourquoi ?* Le registre papier imprimé en 2019 a des cases oui/non ; le formulaire numérique a été conçu plus tard, séparément. — **Arrêt.** Le bureau de l'éducation peut réimprimer le registre ou l'équipe SNIS peut accepter les

deux codages. Les deux se chiffrent en une après-midi.

Aller plus loin — *pourquoi ont-ils été conçus séparément ?* — produit des affirmations vraies sur la coordination institutionnelle sur lesquelles personne dans cette évaluation ne peut agir.

## 7.6 Testez la cause sur les données

Une cause proposée fait une prédiction. Vérifiez-la avant de l'écrire ; cela prend une requête et fait la différence entre un constat et une histoire.

```
# Cause: one school's register uses Y/N. Prediction: Y/N appears in that
# school only, and in most of its months.
yn = attendance[attendance["present"].isin(["Y", "N"])].merge(
    roster[["student_id", "school_id"]], on="student_id", how="left"
)
print(yn["school_id"].value_counts())
print(yn.groupby("school_id")["attendance_date"].nunique())
```

```
attendance |>
  filter(present %in% c("Y", "N")) |>
  left_join(select(roster, student_id, school_id), by = "student_id") |>
  count(school_id)
```

Si les valeurs Y/N s'étaient révélées réparties sur six écoles, la cause ne serait pas un registre — ce serait une convention régionale, et le remède serait différent. **Une cause qui survit à son propre test mérite d'être écrite ; une cause jamais testée est une supposition sur le ton d'une conclusion.**

## 7.7 Placez la cause au niveau où vit le remède

Dernière discipline, et elle décide de l'équité du rapport.

- Une cause de **niveau formation** appelle une action de niveau formation. « La fiche de pointage de FAC020 manque » se corrige en livrant une fiche.
- Une cause de **niveau district** appelle une action de niveau district. « Aucune formation n'a reçu de formulaires en août » ne fait pas trente-huit constats.
- Une cause de **niveau système** appelle une action de niveau système, et la nommer est souvent ce que le rapport contient de plus précieux, même si personne dans la salle ne peut la corriger aujourd'hui.

Attribuer une cause de niveau district aux formations est l'erreur la plus dommageable de la rédaction d'une évaluation. Elle produit un rapportage défensif, les chiffres se lissent, et l'évaluation du trimestre suivant trouve un jeu de données plus propre et moins vrai.

Nommer les causes sert à rendre le tour suivant meilleur. Un rapport qui répartit les responsabilités avec exactitude et ne change rien a échoué à la seule chose pour laquelle il existait.

## 7.8 La suite

Vous avez des constats, et chacun porte désormais une cause avec son niveau et son test. La dernière leçon est le document — ce qu'il contient, dans quel ordre, et les quatre colonnes qui transforment un constat en quelque chose qui sera encore vrai quand quelqu'un vérifiera dans trois mois.

## CHAPITRE 8

# Le rapport qui change le tour suivant

Leçon 8 sur 8 · 60 min

*Constat, cause racine, action corrective, responsable, échéance — cinq colonnes, une ligne par constat. Plus le tour de suivi, seule preuve que tout cela a servi.*

### 8.1 Qui le lit réellement

Trois personnes, et elles lisent trois choses différentes.

- **Le responsable de programme** lit le résumé et veut savoir si les chiffres du rapport du trimestre dernier tiennent.
- **Le point focal SNIS ou suivi-évaluation** lit le tableau des constats et veut savoir ce qu'on lui demande de faire.
- **Le conseiller suivi du bailleur**, six mois plus tard, lit la section méthodologie et veut savoir si vos constats ont été obtenus de façon défendable.

Écrivez pour les trois, dans cet ordre, et la structure en découle.

### 8.2 L'ordre

1. **Ce qui a été évalué.** Jeu de données, période, unités de rapportage dans le champ, indicateurs examinés. Quatre lignes.
2. **Comment.** Quelles dimensions, quels contrôles de bureau, quelles formations visitées et *pourquoi celles-là* — la colonne **reason** de la leçon d'échantillonnage.
3. **Le tableau de bord.** Cinq dimensions, cinq mesures, celle que vous n'avez pas pu mesurer figurant comme constat plutôt que comme case vide.
4. **Les constats.** Le tableau ci-dessous.
5. **Ce que cela implique pour les chiffres publiés.** La section que tout le monde saute et dont le responsable de programme a besoin.
6. **Les limites.** Ce que l'échantillon ne permet pas de soutenir, ce que vous n'avez pas pu vérifier.

La section cinq mérite qu'on y insiste. Une évaluation qui documente des défauts sans dire si le taux de couverture du trimestre dernier doit être réémis a laissé la décision à quelqu'un qui doit maintenant lire les six sections pour la prendre.

### 8.3 Le tableau des constats

Une ligne par constat, et cinq colonnes non négociables.

| # | Constat                                                                                 | Cause racine           |
|---|-----------------------------------------------------------------------------------------|------------------------|
| 1 | Le taux de rapportage tombe à 29 % en août et 45 % en septembre, dans tout le district  | Système — approvis     |
| 2 | SCH09 a sous-déclaré la présence de 42 % (facteur 1,42)                                 | Outil — le registre in |
| 3 | L'équipe d'enquête 2 arrondit les tailles au 0,5 cm (69 % contre 18 à 23 %)             | Capacité — toise lue   |
| 4 | La promptitude ne peut pas être évaluée ; aucune date de transmission dans l'extraction | Outil — date de com    |
| 5 | 24 % des âges tombent sur une année entière exacte (9 % attendus)                       | Méthode — âge estin    |

Trois propriétés font fonctionner ce tableau.

**Chaque constat contient un nombre.** « Le rapportage était mauvais en août » n'est pas un constat ; « 29 %, contre 82 à 92 % les autres mois » en est un. C'est le nombre qui permettra au tour de suivi de dire si quelque chose a bougé.

**Chaque responsable est une fonction, et une personne l'occupe.** « Le district » ne se relance pas. Écrivez la fonction, et vérifiez que quelqu'un l'occupe effectivement — une action confiée à un poste vacant est un constat en soi.

**Chaque échéance est une date.** « En continu » et « dès que possible » signifient tous deux jamais, et tout le monde dans la salle le sait quand c'est écrit.

## 8.4 Des constats qu'un lecteur peut vérifier

Rédigez chacun de sorte que quelqu'un disposant des mêmes données puisse le reproduire. C'est ce qui sépare une évaluation d'une opinion, et cela coûte une proposition subordonnée.

```
findings = pd.DataFrame([
    {
        "id": 2,
        "dimension": "Accuracy",
        "unit": "SCH09",
        "measure": "verification factor",
        "value": 1.416,
        "comparator": "district 1.012, 23 of 24 schools exactly 1.000",
        "evidence": "outputs/verification-factors.csv",
    },
])
findings.to_csv("outputs/dqa-findings.csv", index=False)

findings <- tibble::tribble(
  ~id, ~dimension, ~unit, ~measure, ~value, ~comparator,
  2L, "Accuracy", "SCH09", "verification factor", 1.416, "district 1.012"
)

readr::write_csv(findings, here::here("outputs", "dqa-findings.csv"))
```

La colonne `comparator` est celle qu'on omet. Un facteur de vérification de 1,42 ne veut rien dire tant que le lecteur ignore que les vingt-trois autres écoles étaient à 1,000 — c'est cette comparaison qui transforme un nombre en constat.

La colonne `evidence` désigne le fichier qui l'a produit, même discipline que le journal de nettoyage — **l'artefact voyage avec l'affirmation**.

## 8.5 Notez les dimensions, ne notez pas le district

Publiez le tableau de bord comme cinq mesures avec leurs sources. Si un composite est exigé, placez-le à côté et énoncez les pondérations.

```
scorecard = pd.DataFrame({
    "dimension": ["Completeness", "Timeliness", "Accuracy", "Consistency", "Integrity"],
    "measure": ["Reporting rate", "Not assessable", "Verification factor",
               "Outlier rate", "Round-number share"],
    "value": ["76.5%", "-", "1.012 district / 1.42 worst unit",
             "11 of 2,094 below half own median", "10.4%, chance level"],
    "finding": [1, 4, 2, None, None],
})
```

```
scorecard <- tibble::tribble(
  ~dimension, ~measure, ~value, ~finding,
  "Completeness", "Reporting rate", "76.5%", 1L,
  "Timeliness", "Not assessable", "-", 4L,
  "Accuracy", "Verification factor", "1.012 district / 1.42 max", 2L,
  "Consistency", "Outlier rate", "11 of 2,094", NA_integer_,
  "Integrity", "Round-number share", "10.4%, chance level", NA_integer_
)
```

Remarquez que la ligne d'exactitude porte deux nombres. **Une dimension résumée par un seul nombre masque la distribution**, et la distribution était tout le constat.

Remarquez aussi les deux lignes sans constat rattaché. Intégrité et cohérence sont revenues propres, et le dire explicitement vaut autant que les constats — le lecteur sait quels contrôles ont été exécutés et passés, au lieu de se demander s'ils l'ont été.

## 8.6 Le tour de suivi est l'essentiel

Une évaluation ne vaut rien à elle seule. Elle prend de la valeur quand les mêmes contrôles tournent à nouveau et que les nombres bougent.

```
history = pd.concat([
    pd.read_csv(p).assign(round=p.stem)
    for p in sorted(Path("outputs/dqa").glob("*.csv"))
])
print(history.pivot_table(index="measure", columns="round", values="value"))

history <- purrr::map_dfr(
  list.files(here::here("outputs", "dqa"), full.names = TRUE),
  ~ readr::read_csv(.x) |> dplyr::mutate(round = tools::file_path_sans_ext(basename(.x)))
)

tidyr::pivot_wider(history, id_cols = measure, names_from = round, values_from = value)
```

| Mesure                      | T3       | T4       | Évolution |
|-----------------------------|----------|----------|-----------|
| Taux de rapportage          | 76,5 %   | 88,1 %   | +11,6 pt  |
| Écoles hors tolérance       | 1 sur 24 | 0 sur 24 | clos      |
| Tailles arrondies, équipe 2 | 69 %     | 24 %     | clos      |

Ce tableau est le livrable qui justifie le travail. Il donne aussi un verdict à chaque action corrective — faite, non faite, ou faite sans effet — ce qui est exactement la redevabilité que demande l'engagement de la CHS sur la gestion de l'information.

**Reportez chaque action ouverte.** Une action qui figure dans trois rapports consécutifs sans se clore est elle-même un constat, et généralement de niveau système.

## 8.7 La présenter à ceux qu'elle décrit

Dernier point, et il décide si le tour suivant sera meilleur ou plus silencieux.

- **Montrez-la-leur d'abord.** Un constat qu'une formation découvre dans un rapport bailleur sera contesté, et la contestation portera sur votre méthode plutôt que sur les données.
- **Commencez par les constats système.** Ouvrir sur l'approvisionnement en formulaires d'août établit que l'évaluation regarde toute la chaîne et non des individus.
- **Nommez ce qui était propre.** Vingt-trois écoles sur vingt-quatre étaient exactes. Dites-le avant de dire que la vingt-quatrième ne l'était pas.
- **Demandez-leur la cause.** Ceux qui ont rempli le registre savent en général pourquoi il a cette allure, et ils vous le diront si la réunion n'est pas un tribunal.

La mesure d'une évaluation de la qualité des données n'est pas le nombre de constats produits. C'est de savoir si les données du trimestre suivant sont meilleures, et le rapportage défensif est le seul résultat qui garantit qu'elles ne le seront pas.

## 8.8 Ce que ce cours, et ce module, vous laissent

Vous savez profiler un export neuf et consigner chacune de vos modifications. Vous savez assembler plusieurs fichiers et prouver que les jointures ont fait ce que vous aviez annoncé. Et vous savez désormais évaluer le résultat comme le fera un auditeur — cinq dimensions avec leurs mesures, un recomptage face à la source, un échantillon défendable, et un rapport où chaque constat a un responsable et une date.

Le module 2 est complet, et c'est celui que tous les cours sectoriels supposent acquis.

Le module 3, *Indicateurs et mesure*, est ce à quoi tout cela préparait. Il s'ouvre sur *Conception d'indicateurs et cadre logique* — rédiger des définitions assez précises pour que deux analystes les calculant séparément obtiennent le même chiffre, ce qui est le défaut sous-jacent à plus de la moitié des constats de ce cours.

## À propos de cette édition

Ce PDF est produit à partir de la même source que la version web de la formation ; les deux ne peuvent donc pas diverger. L'édition en ligne comporte le code exécutable, les jeux de données téléchargeables et les corrections publiées depuis la production de ce fichier.

Tous les jeux de données cités dans cette formation sont synthétiques ou entièrement anonymisés. Aucun enregistrement ne renvoie à une personne réelle.

Cette formation est publiée en anglais et en français. L'édition française est une adaptation professionnelle reprenant la terminologie employée par l'UNICEF, l'OMS, le PAM, OCHA et les ministères nationaux — et non une traduction littérale.

Consulter et exécuter la formation en ligne sur [data-analysis.cassion.dev/fr/cours/data-quality-assessment](https://data-analysis.cassion.dev/fr/cours/data-quality-assessment)

Le contenu pédagogique est diffusé sous licence CC BY 4.0. Vous pouvez l'adapter et le rediffuser, y compris à des fins commerciales, en créditant Cassion.

Généré le 28 juillet 2026.